Retrieval of atmospheric CH4 profiles from Fourier transform infrared data using dimension reduction and MCMC

Tukiainen, S.; Railo, J.; Laine, M.; Hakkarainen, J.; Kivi, R.; Heikkinen, P.; Chen, H.; Tamminen, J.

Published in:
Journal of geophysical research-Atmospheres

DOI:
10.1002/2015JD024657

AGU PUBLICATIONS

JGR

Journal of Geophysical Research: Atmospheres

RESEARCH ARTICLE
10.1002/2015JD024657

Retrieval of atmospheric CH_4 profiles from Fourier transform infrared data using dimension reduction and MCMC

S. Tukiainen[1], J. Railo[2], M. Laine[1], J. Hakkarainen[1], R. Kivi[3], P. Heikkinen[3], H. Chen[4,5], and J. Tamminen[1]

[1]Earth Observation Unit, Finnish Meteorological Institute, Helsinki, Finland, [2]Department of Mathematics and Statistics, University of Jyväskylä, Jyväskylä, Finland, [3]Arctic Research Center, Finnish Meteorological Institute, Sodankylä, Finland, [4]Centre for Isotope Research (CIO), Energy and Sustainability Research Institute Groningen (ESRIG), University of Groningen, Groningen, Netherlands, [5]Cooperative Institute for Research in Environmental Sciences (CIRES), University of Colorado Boulder, Boulder, Colorado, USA

Key Points:
- Retrieval of CH_4 profile information from FTIR data, potentially improving the total column
- Practical and intuitive dimension reduction methodology
- Full Bayesian uncertainty quantification using Markov chain Monte Carlo

Correspondence to:
S. Tukiainen,
simo.tukiainen@fmi.fi

Citation:
Tukiainen, S., J. Railo, M. Laine, J. Hakkarainen, R. Kivi, P. Heikkinen, H. Chen, and J. Tamminen (2016), Retrieval of atmospheric CH_4 profiles from Fourier transform infrared data using dimension reduction and MCMC, *J. Geophys. Res. Atmos.*, *121*, 10,312–10,327, doi:10.1002/2015JD024657.

Received 15 DEC 2015
Accepted 11 AUG 2016
Accepted article online 16 AUG 2016
Published online 1 SEP 2016

Abstract We introduce an inversion method that uses dimension reduction for the retrieval of atmospheric methane (CH_4) profiles. Uncertainty analysis is performed using the Markov chain Monte Carlo (MCMC) statistical estimation. These techniques are used to retrieve CH_4 profiles from the ground-based spectral measurements by the Fourier Transform Spectrometer (FTS) instrument at Sodankylä (67.4°N, 26.6°E), Northern Finland. The Sodankylä FTS is part of the Total Carbon Column Observing Network (TCCON), a global network that observes solar spectra in near-infrared wavelengths. The high spectral resolution of the data provides approximately 3 degrees of freedom about the vertical structure of CH_4 between around 0 and 40 km. We reduce the dimension of the inverse problem by using principal component analysis. Smooth and realistic profiles are sought by estimating three parameters for the profile shape. We use Bayesian framework with adaptive MCMC to better characterize the full posterior distribution of the solution and uncertainties related to the retrieval. The retrieved profiles are validated against in situ AirCore soundings which provide an accurate reference up to 20–30 km. The method is presented in a general form, so that it can easily be adapted for other applications, such as different trace gases or satellite-borne measurements where more accurate profile information and better analysis of the uncertainties would be highly valuable.

1. Introduction

Greenhouse gas measurements of the atmosphere are necessary for monitoring and understanding the observed global warming of the Earth. Greenhouse gases such as water vapor (H_2O), carbon dioxide (CO_2), and methane (CH_4) absorb electromagnetic radiation, originally emitted from the Sun—and reradiated from the Earth—directly affecting the energy balance of the Earth. The theory of this radiative impact is well established and it has been recently confirmed by direct measurements as well [*Feldman et al.*, 2015]. The current understanding is that the increased greenhouse gas concentrations have warmed the globe approximately by 0.85 K since the end of the 1800 century [*Intergovernmental Panel on Climate Change*, 2013, section 2.]. Because greenhouse gases are important factors in the behavior of the atmosphere, accurate estimates of the abundances and trends of the greenhouse gases are crucial constraints for meaningful climate model simulations of the future.

Greenhouse gas concentrations are traditionally measured by analyzing local air samples or by using ground-based Fourier transform spectrometers (FTS) that use Fourier transform infrared (FTIR) spectroscopy to measure column-integrated values. Examples of the FTS instrumentation include the Total Carbon Column Observing Network (TCCON) network [*Wunch et al.*, 2011] which consists of around 20 measurement sites around the world and the Network for the Detection of Atmospheric Composition Change (NDACC) which has ~23 stations. More recently, satellite-based measurements have been gathered as well. The SCIAMACHY instrument [*Gottwald and Bovensmann*, 2011] on board the Envisat satellite (2002–2012) was one of the first satellite equipments to continuously measure atmospheric greenhouse gases on a global scale. The SCIAMACHY data were also used to derive one of the first greenhouse gas trends from space [*Frankenberg et al.*, 2011; *Schneising et al.*, 2011] even though the data became noisier in the last years of the mission. SCIAMACHY measured greenhouse gases using nadir geometry and near-infrared/shortwave infrared

(NIR/SWIR) wavelengths. The same concept have been used by the Greenhouse gases Observing SATellite (GOSAT), launched in 2009, and NASA's Orbiting Carbon Observatory 2 (OCO 2), launched in 2014. GOSAT was the first dedicated satellite mission to measure CO_2 and CH_4 (and O_2 in the oxygen A band), providing about 500 clear sky measurements each day. These measurements are used in various studies of the biosphere, such as in investigations of the water vapor and chlorophyll fluorescence. OCO 2 provides an order of magnitude more CO_2 data than GOSAT, and 2 orders more than SCIAMACHY. The first versions of the OCO 2 level 2 data are already released and under investigation by the scientific community. Several other instruments have measured greenhouse gases from nadir using thermal infrared (TIR) wavelengths. These instruments include, for example, AIRS on the Aqua satellite [*Aumann et al.*, 2003], TES on Aura [*Beer*, 2006], and IASI on Metop A/B [*Aires et al.*, 2002]. MIPAS on Envisat [*Fischer et al.*, 2008] was a limb-viewing instrument using TIR, and ACE-FTS instrument on SCISAT-1 [*Bernath et al.*, 2005] uses solar occultation, covering wavelengths from SWIR to TIR. ACE-FTS provides also CH_4 profile data with relatively good vertical resolution (3–4 km).

The ground-based measurement networks are crucial for validating the satellite-based observations. Especially the TCCON network has been widely used as the reference [e.g., *Dils et al.*, 2014]. The TCCON instruments look directly at the Sun providing good signal-to-noise ratio in clear sky conditions. Furthermore, the TCCON instruments achieve very good spectral resolution (typically ~0.02 cm^{-1}). The TCCON total column results are delivered as column-averaged dry air mole fractions (denoted by XCO_2, XCH_4, etc.), i.e., the gas columns are divided by the dry air column which is approximated from the retrieved O_2 column. This is a standard practice to get a more stable quantity, less affected by variations in the air pressure and water vapor. Using the retrieved O_2 as a reference is also useful because it removes some of the systematic errors related to the retrieval and compensates some of the instrumental errors. The same concept is used with the satellite-based columns as well. The accuracy requirements for the retrieved products are exceptionally strict. For example, the average CO_2 column should be retrieved within ~1 ppm (out of ~400 ppm) and the CH_4 column within ~5 ppb (out of ~1800 ppb). These are difficult limits to achieve with remote sensing measurements. Although accurate measurement data with high signal-to-noise ratio and spectral resolution are necessary, accurate forward modeling and inversion methods are crucial as well.

The radiative transfer in the shortwave infrared (SWIR) and near-infrared (NIR) wavelengths has been extensively studied in the past. There are more than 20 published SWIR/NIR radiative transfer codes available such as MODTRAN, DISORT, 4A/OP, MOSART and SCIATRAN. The fundamental theory behind the absorption line formation and the line shape are covered in numerous textbooks such as *Goody and Yung* [1995] and *Liou* [2002]. The line parameters have been derived using laboratory measurements and there are line databases available for a large number of isotopologues and wavelength ranges [e.g., *Rothman et al.*, 2013; *Jacquinet-Husson et al.*, 2011].

Modern retrieval methods are usually based on the Bayesian statistical approach, which exploits the prior information about the state of the atmosphere [e.g., *Butz et al.*, 2011]. A posterior analysis is then performed by combining the prior and the likelihood. A canonical example is the optimal estimation method by *Rodgers* [2000]. Examples of retrieval methods for satellite-based greenhouse gas measurements are described in *Buchwitz et al.* [2006] for SCIAMACHY, *Yoshida et al.* [2011] for GOSAT, and in *O'Dell et al.* [2012] for OCO 2. The official TCCON retrieval uses the GGG software package [*Wunch et al.*, 2011], which solves the inversion problem by scaling the prior profile densities of the gases that absorb in the wavelength window of interest. Several wavelength windows are used for most of the gases and the final estimate of the profile is an error-weighted average of the individual results. The absorption coefficients are evaluated beforehand, assuming known fixed temperature and pressure profiles.

Scaling of a prior profile, the method used in GGG, is an intuitive and robust concept but retrieves only one piece of information and assumes that the shape of the prior profile is known. There have been efforts to retrieve also profile information from FTIR data. In *Kuai et al.* [2012], the CO_2 prior profile is scaled on three levels instead of having only one common scaling factor. Furthermore, *Kuai et al.* [2013] assimilate FTIR data and thermal infrared satellite data to estimate lower tropospheric CO_2 columns, and *Hase et al.* [2004] report optimal estimation-based profile retrievals from high-resolution FTIR measurements. In *Senten et al.* [2012], the authors use information operator approach (IOA) [*Doicu et al.*, 2007] to retrieve profiles from FTIR data and investigate other common methods as well.

In this paper we describe an alternative inversion method for retrieving greenhouse gas profiles from FTIR measurements. We demonstrate a simple and flexible technique to retrieve information about the vertical

structure of the atmosphere, leading to a more accurate estimate of the total column. The retrieval method presented is based on reducing dimension of the state space by a suitable truncation of the prior covariance matrix, hence reducing the dimension of the problem. For the estimation of the unknown parameters we use adaptive Markov chain Monte Carlo (MCMC) methods. The MCMC solution gives a better understanding of the posterior uncertainties, compared to the derivative-based estimation methods. Once the natural variability of the retrieved gas is properly characterized, our retrieval method is only marginally affected by the possibly incorrect prior profile shape. In this work we retrieve CH_4 profiles, but the same technique can be used with other trace gases as well.

2. Sodankylä Measurements

The Sodankylä FTS station is part of the Finnish Meteorological Institute's Arctic Research Center in Northern Finland (67.3668°N, 26.6310°E). The Sodankylä FTS [*Kivi and Heikkinen*, 2016] has been operational since February 2009, providing direct Sun measurements from February to November. Up to several hundred measurements a day are recorded depending on the season and cloudiness, but the wintertime measurements are not possible due to the absence of the sunlight. The Sodankylä FTS is a Bruker IFS 125 HR with a A547N solar tracker. It has three detectors: InGaAs (4000–11,000 cm^{-1}), Si (11,000–15,000 cm^{-1}), and InSb (1800–6000 cm^{-1}). The wave number limits are roughly the useful ranges of the detectors. The Sodankylä FTS operates on the optical path difference of 45 cm, with the 2.3923 mrad field of view, leading to the spectral resolution of ~0.02 cm^{-1}.

The measurements are routinely processed using the GGG software and distributed as column-averaged dry-air mole fractions. The latest version of the algorithm, GGG2014 [*Wunch et al.*, 2015], retrieves CH_4 from three separate wavelength windows centered at 5938, 6002, and 6076 cm^{-1}. In general, the CH_4 errors within the GGG2014 retrieval are below 0.5% (about 5 ppb) for solar zenith angles less than 85° [*Wunch et al.*, 2015]. One of the main sources of error comes from the prior profile shape. Especially the springtime prior profiles of CH_4 are prone to differ substantially from the truth. The prior profiles for Sodankylä depend on the tropopause height but not on the polar vortex which can largely affect the stratospheric part of the CH_4 profile.

In addition to the FTS measurements, a variety of in situ observations, e.g., balloon soundings and tower measurements, are collected at the Sodankylä site. To validate CH_4 profiles retrieved from the Sodankylä FTIR data, we use the AirCore profile measurements made along with the FTIR measurements. AirCore is an innovative atmospheric sampling tool made of a long coil of tubing which slowly fills with ambient air during the payload descent of the balloon sounding [*Karion et al.*, 2010]. The vertical profile measurements used in this study were made by an ~100 m long AirCore and obtained by the analysis of the air samples on the ground using a cavity ring-down spectrometer (Picarro Inc. model G2401) within a few hours after landing. In this study we use Sodankylä AirCore version 1.0 data. The AirCore soundings made at Sodankylä are important because there are not much vertically resolved in situ data of the Arctic methane, especially measured in the stratospheric polar vortex conditions.

3. Forward Model

The radiative transfer problem of the FTS measurement is solved here with the Matlab code SWIRLAB (https://github.com/tukiains/swirlab) which can be used to calculate absorption coefficients, model short-wave infrared radiative transfer, and perform retrievals from measurement data. So far we have tested SWIRLAB with the ground-based geometry where the instrument is looking directly at the Sun. In this simplified case, the scattering of radiation from air molecules and aerosols is negligible and only the absorption of the atmosphere needs to be considered. The satellite viewing geometry is not utilized in this work.

In this study we assume a spherical Earth and atmosphere. The SWIRLAB atmosphere can be discretized with an arbitrary layering but we have mostly used 100 layers between 0 and 70 km. The layering is based on a geometric series, where the layers are thinnest close to the surface (~800 m) and thickest at the top (~1.25 km). This kind of layering is rational as the number density of methane decreases exponentially with altitude, although in our tests the retrieval results are not very sensitive to the layering scheme. The layers are homogeneous, i.e., the number density, temperature, and pressure inside each layer are assumed constant. All the layer parameters (densities, temperature, and pressure) are linearly interpolated to the center of each layer.

For the calculation of the absorption coefficients we use the HITRAN2012 line database [*Rothman et al.*, 2013]. The line intensities and positions are computed according to *Rothman et al.* [1998] with the total internal partition sums from *Laraia et al.* [2011]. Isotopologue ratios are taken from the HITRAN2012 database and kept fixed in the retrieval. Line mixing is not taken into account. For the line shape calculation we use the Voigt profile, which is generally a function of temperature and pressure (and marginally the partial pressure of the absorbing gas). In our retrieval, we use the same temperature, pressure, and prior profiles for the trace gases as GGG2014. The solar spectrum is also the same. The GGG2014 prior profiles are generated using National Centers for Environmental Prediction/National Center for Atmospheric Research model analysis data (for temperature, pressure. and humidity), empirical models for the greenhouse gases, and a variety of satellite-based and in situ data [*Wunch et al.*, 2011].

We note that our absorption coefficient calculation is somewhat simplified. For example, in addition to the HITRAN database, the OCO CO_2 retrieval employs results from *Devi et al.* [2007], *Predoi-Cross et al.* [2009], and *Sung et al.* [2009], just for the 1.61 μm band alone. Line mixing, independent isotopologues, and water vapor content of the atmosphere should be taken into account for the best possible results. Also, in some cases, especially for the O_2 retrievals with high solar zenith angles, the Voigt profile might not be the best approximation for the line shape. Nevertheless, our simple forward modeling is enough to adequately model the FTIR data and demonstrate the benefits of our retrieval method.

4. CH_4 Profile Retrieval

Methane is an important carbon-containing species in the atmosphere. Atmospheric methane is produced by anthropogenic and natural processes at the surface; it has no source in the atmosphere. Methane has a relatively long lifetime (8–10 years) allowing it to be transported from the troposphere to the stratosphere. In the troposphere, CH_4 has a fairly constant mole fraction, but in the stratosphere, CH_4 is an excellent tracer to study transport processes [*Brasseur and Solomon*, 2005]. Thus, it is useful to measure the vertical distribution of CH_4. Besides, an accurate vertical profile gives a credible estimate of the total column, which often is an important variable.

Ground-based FTS instruments measure solar light modified by the whole atmospheric column of CH_4 and other interfering trace gases. Our retrieval problem is to estimate the vertical CH_4 profile from a single spectrum. The information about the vertical structure comes from the pressure and temperature dependence of the Voigt line shape: absorption lines are wider close to the Earth in high pressure than higher in the atmosphere. Given the measured spectrum $\boldsymbol{y} \in \mathbb{R}^m$, where m is the number of wavelengths, we wish to approximate the state vector $\boldsymbol{x} \in \mathbb{R}^n$, where n is the number of atmospheric layers times the number of trace gases to be retrieved. Typically, the number of layers is around 50 or more depending on the forward model. Thus, the state vector $\boldsymbol{x}$ represents the densities of the discretized atmosphere in the forward model, and the inverse problem can be formulated in a very general way

$$\hat{\boldsymbol{x}} = \mathbf{R}(\boldsymbol{y}, \boldsymbol{C}_y, \theta), \tag{1}$$

where $\hat{\boldsymbol{x}} \in \mathbb{R}^n$ is the retrieved state, $\mathbf{R}$ is the retrieval method, and $\boldsymbol{C}_y \in \mathbb{R}^{m \times m}$ is the measurement error covariance. Typically, the problem is also affected by some additional parameters, denoted by θ, that can be either fixed or retrieved. If the extra parameters are retrieved, as they often are, they should be part of $\hat{\boldsymbol{x}}$. Nevertheless, we ignore them in the following to keep the equations simpler to follow. The fundamental problem in the estimation of the vertical profile $\hat{\boldsymbol{x}}$ from a single spectral measurement is that the measured noisy spectrum does not contain enough information to independently resolve all n altitude levels in a meaningful way (the inverse problem is said to be *ill-posed*). In other words, without proper prior regularization, there are a vast number of profiles that fit the data but are physically unrealistic, e.g., are oscillating or otherwise unstable.

In order to make the problem well posed, we need to regularize, or constrain, the retrieval by using additional information about the solution. For example, we can require a certain degree of smoothness from the retrieved profile or demand the retrieved densities to be positive. A general strategy is to cast the inversion problem in statistical form and consider unknowns and other auxiliary information as probability distributions. In Bayesian statistical approach we assume a priori information about the state vector $\boldsymbol{x}$ and use the Bayes' theorem to combine information from the prior measurement in the estimation process. In general probability distribution form, this can be written as

$$\boldsymbol{p}(\boldsymbol{x}|\boldsymbol{y}) \propto \boldsymbol{p}(\boldsymbol{y}|\boldsymbol{x})\boldsymbol{p}(\boldsymbol{x}), \tag{2}$$

where $\boldsymbol{p}(\boldsymbol{x}|\boldsymbol{y})$ is the posterior distribution of the state $\boldsymbol{x}$ given the observations $\boldsymbol{y}$, $\boldsymbol{p}(\boldsymbol{y}|\boldsymbol{x})$ is the likelihood and $\boldsymbol{p}(\boldsymbol{x})$ is the prior distribution. In our profile retrieval, the prior is a CH_4 profile $\boldsymbol{x}_0 \in \mathbb{R}^n$ having the prior error covariance $\boldsymbol{C}_x \in \mathbb{R}^{n\times n}$. Assuming that the prior distribution and the distribution defining the likelihood are independent Gaussian, then the posterior density can be written as

$$\boldsymbol{p}(\boldsymbol{x}|\boldsymbol{y}) \propto \exp\left(-\frac{1}{2}\left(||\boldsymbol{y} - f(\boldsymbol{x})||^2_{\boldsymbol{C}_y} + ||\boldsymbol{x} - \boldsymbol{x}_0||^2_{\boldsymbol{C}_x}\right)\right), \tag{3}$$

where we use the notation $||\boldsymbol{b}||^2_{\boldsymbol{A}} = \boldsymbol{b}^T\boldsymbol{A}^{-1}\boldsymbol{b}$, and $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ is the forward model. The Bayesian inversion is, in principle, an elegant way to assimilate prior model information about $\boldsymbol{x}$ and the information in the data $\boldsymbol{y}$. However, in many cases the prior is mainly used to produce a "well-behaving" posterior, e.g., the desired smoothness in the retrieved profile, instead of a careful statistical analysis of the prior as the information on the natural variability in the profiles and the smoothing properties of the measurement system. In the FTIR inverse problem, the posterior depends significantly on the prior mean profile $\boldsymbol{x}_0$ and the relationship between $\boldsymbol{C}_y$ and $\boldsymbol{C}_x$. Too loose prior distribution will lead to an unstable solution, and too tight will bias the result if $\boldsymbol{x}_0$ is far from the truth. Nevertheless, the vertical variability allowed by the prior is usually set relatively restricted to substantially regularize the solution and to avoid any superfluous oscillation in the retrieved profile. To make the uncertainty quantification from the posterior distribution valid, the prior should be an honest statistical representation of the model knowledge before the measurement is taken, but it is impractical to use such a prior when the state vector contains many more unknowns than there are degrees of freedom in the signal.

4.1. Dimension Reduction Method

In mathematical sense, a vertical profile is a function and as such an infinite-dimensional object. When we discretize the profile with, say, 70 levels, the corresponding retrieval problem has 70 unknowns. The intrinsic dimension of the problem is typically much lower, as the observations contain only limited amount of vertical information. This dimension is a property of the measurement system, and together with the prior constraints posed for the retrieval process, it determines the resolution of the retrieved profile as manifested in the averaging kernel, see section 4.4 later in this paper or, e.g., *Rodgers* [2000, section 2.4]. The approach presented in this paper is based on the work by *Marzouk and Najm* [2009] and *Solonen et al.* [2016]. In atmospheric remote sensing, the dimension reduction techniques have been utilized, e.g., by *Masiello et al.* [2012] and *Cui et al.* [2014].

If we want to solve the problem in the full dimension determined by the discretization, we need to use some prior constraints due to the ill-posed nature of the problem. In addition, using the full dimension would make the corresponding estimation problem harder to solve. In this work, we utilize sampling-based uncertainty analysis, which allows the use of the full Bayesian posterior distribution by MCMC simulation. For MCMC, the computational efficiency is heavily affected by the dimension of the unknown and dimension reduction will be of great benefit as the use of a low-rank approximation of the prior covariance matrix will efficiently restrict the MCMC sampling to a lower dimensional space.

In dimension reduction, we formulate the problem as a low-dimensional problem that can be mapped back to the original dimension determined by the discretization, which will make the method discretization invariant. If there are only 3 degrees of freedom in the observations, then we can parameterize the problem so that we have only three parameters to estimate, but this parameterization must be chosen so that the information available is retained. In the method presented here, the prior distribution that defines the dimension reduction is selected so that it reflects the information content of the observations. An extreme case would be just to scale the prior profile, $\boldsymbol{x} = \alpha\boldsymbol{x}_0$, i.e., to have a one-dimensional problem as in the operational TCCON inversion algorithm.

Let us consider a parameterization for the profile $\boldsymbol{x}$ in terms of a parameter vector $\boldsymbol{\alpha}$ and a projection matrix $\boldsymbol{P}$ as

$$\boldsymbol{x} = \boldsymbol{x}_0 + \boldsymbol{P}\boldsymbol{\alpha}, \tag{4}$$

where $\boldsymbol{x}_0 \in \mathbb{R}^n$ is a mean profile. Our aim is to define $\boldsymbol{P}$ in such a way that it defines a probability distribution of $\boldsymbol{x}$, with $\boldsymbol{\alpha}$ as simple as possible. If the prior distribution for $\boldsymbol{x}$ is Gaussian, $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{x}_0, \boldsymbol{C})$, where $\boldsymbol{C}$ is an $n \times n$ positive definite prior covariance matrix, then we can choose $\boldsymbol{P}$ such that $\boldsymbol{P}\boldsymbol{P}^T = \boldsymbol{C}$ and $\boldsymbol{\alpha} \sim \mathcal{N}(0, \boldsymbol{I}_n)$. Then it follows that

$$\mathrm{cov}(\boldsymbol{x}) = \mathrm{cov}(\boldsymbol{P}\boldsymbol{\alpha}) = \boldsymbol{P}\mathrm{cov}(\boldsymbol{\alpha})\boldsymbol{P}^T = \boldsymbol{P}\boldsymbol{I}_n\boldsymbol{P}^T = \boldsymbol{C}. \tag{5}$$

Next, to define the dimension reduction, we will use a similar parameterization, but replace the prior covariance matrix $\boldsymbol{C}$ by its low-rank version $\widetilde{\boldsymbol{C}}$ and use it to build a projection from a lower-dimensional subspace to the original n-dimensional space. We factorize the covariance $\boldsymbol{C}$ using the singular value decomposition,

$$\boldsymbol{C} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T = \sum_{i=1}^{n} \lambda_i \boldsymbol{u}_i \boldsymbol{u}_i^T, \tag{6}$$

with

$$\boldsymbol{U} = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n], \quad \boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n), \tag{7}$$

where $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n$ are the singular vectors, and $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n > 0$ are the singular values (or, equivalently, the eigenvalues) of $\boldsymbol{C}$. The reduced rank version of $\boldsymbol{C}$ of order k, $1 \leq k < n$, is defined as

$$\widetilde{\boldsymbol{C}} = \sum_{i=1}^{k} \lambda_i \boldsymbol{u}_i \boldsymbol{u}_i^T = \boldsymbol{P}_k \boldsymbol{P}_k^T, \tag{8}$$

where $\boldsymbol{P}_k = [\sqrt{\lambda_1}\boldsymbol{u}_1, \ldots, \sqrt{\lambda_k}\boldsymbol{u}_k] \in \mathbb{R}^{n \times k}, k < n$, is a projection matrix from $\mathbb{R}^k$ to $\mathbb{R}^n$. The reduced matrix $\widetilde{\boldsymbol{C}} \in \mathbb{R}^{n \times n}$ approximates the original covariance $\boldsymbol{C}$ and it is the best rank k approximation according to the Eckart-Young-Mirsky theorem of numerical linear algebra.

With the reduced rank covariance matrix $\widetilde{\boldsymbol{C}} = \boldsymbol{P}_k \boldsymbol{P}_k^T$, we can define a low-dimensional parameterization for the profile

$$\boldsymbol{x} = \boldsymbol{x}_0 + \boldsymbol{P}_k \boldsymbol{\alpha}_k, \tag{9}$$

where $\boldsymbol{\alpha}_k$ is a k-dimensional vector whose prior distribution will be a k-dimensional Gaussian $\mathcal{N}(0, \boldsymbol{I}_k)$. The optimizer for the parameter estimation problem can now work in the reduced dimension, while the forward model is still run in the full dimension. With a sampling-based approach, such as MCMC, we can make proposal draws for $\boldsymbol{\alpha}_k$ from the reduced dimension, and then generate full profile using equation (9). If the neglected singular values are small, i.e., whose singular vectors contribute to the variability (and thus to the uncertainty) less than a selected threshold, the original prior information will be represented accurately.

In practice, it is useful to estimate the log profile instead of the original,

$$\log(\boldsymbol{x}) = \log(\boldsymbol{x}_0) + \boldsymbol{P}_k \boldsymbol{\alpha}_k, \tag{10}$$

as this will always lead to positive profiles, which is not the case if we define the estimation problem with the Gaussian prior of equation (9). This means that the prior for $\boldsymbol{x}$ is lognormal, and we have to define the prior covariance matrix $\boldsymbol{C}$ in terms of the log profiles.

In the CH_4 profile retrieval using FTIR data, we define the posterior distribution in terms of the $\boldsymbol{\alpha}$ and retrieve k parameters instead of n. The posterior density of equation (3) now becomes

$$\boldsymbol{p}(\boldsymbol{\alpha}|\boldsymbol{y}) \propto \exp\left(-\frac{1}{2}\left(||\boldsymbol{y} - f(\boldsymbol{x})||^2_{\boldsymbol{C}_y} + ||\boldsymbol{\alpha}||^2_{\boldsymbol{I}_k}\right)\right), \tag{11}$$

where $\boldsymbol{y} \in \mathbb{R}^m$ is the measurement vector, $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ is the forward model, and $\boldsymbol{C}_y \in \mathbb{R}^{m \times m}$ is the measurement error covariance. As already mentioned, the prior covariance of the $\boldsymbol{\alpha}$ parameters is $\boldsymbol{I}_k \in \mathbb{R}^{k \times k}$, i.e., a diagonal unit matrix. While equations (3) and (11) are analogous, the difference is that in the dimension reduction retrieval the likelihood term receives only smooth realization of $\boldsymbol{x}$, as they are generated by equation (9) or equation (10), but in equation (3) the state vector has no such restriction. Thus, the estimation of the posterior with k parameters in the state vector is numerically much more stable than with n parameters.

4.2. Prior Covariance

The important part of the dimension reduction based retrieval is the prior covariance that has to be selected before the estimation. In this work we construct the prior covariance from general assumptions, but we do use ACE-FTS satellite measurements to check that our assumptions are valid. For CH_4, we allow some variability between 0 and ~10 km, a large variability in the UTLS region (upper troposphere, lower stratosphere) between ~10 and ~35 km, and very little variability above ~35 km. From the UTLS we expect to find the largest gradients in the CH_4 mole fraction, a well-known feature visible also in the ACE-FTS version 2.2 data (Figure 1). Note that the ACE-FTS data set we use may overestimate the natural variation because it contains seasonal and year-to-year variations. The altitudes above ~35 km are "fixed" to the prior profile because the measurement has very little sensitivity to these altitudes. Below the (approximate) tropopause altitude,

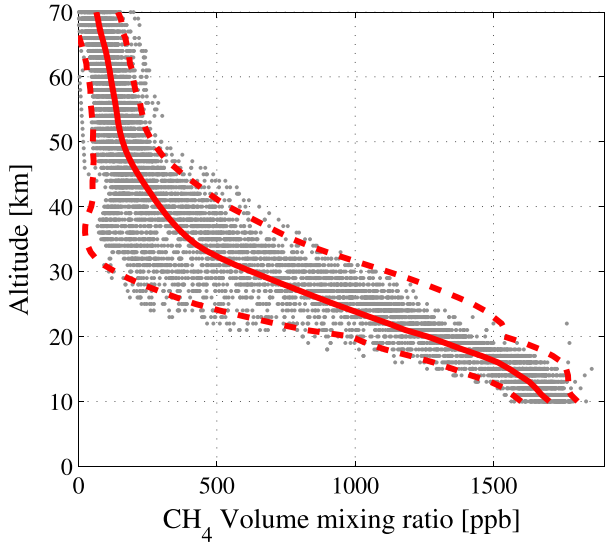

Figure 1. CH_4 profiles (grey), mean (solid line), and $\pm 2\sigma$ (dashed lines) over Sodankylä measured by ACE-FTS in 2004–2010.

we assume that CH_4 is rather well mixed and the prior profile has approximately the correct shape at least. In a recent CH_4 retrieval paper by *Sepúlveda et al.* [2014] the authors describe typical CH_4 signals between 0 and 25 km and have a similar treatment of the variability than we have. However, they also consider possible CH_4 enhancement on the ground level, a feature that we do not take into account.

To set up the prior covariance matrix, we first define its diagonal. The prior standard deviations for each altitude h are calculated using a mixture of two square exponential "bumps" as

$$\sigma(h) = \sigma_1 \exp\left(-(h-h_1)^2 s_1^{-2}\right) + \sigma_2 \exp\left(-(h-h_2)^2 s_2^{-2}\right). \tag{12}$$

The terms in equation (12) are chosen so that the diagonal elements between 10 and 30 km approximately resemble the natural variability of CH_4 over Sodankylä, measured by ACE-FTS in 2004–2010. We define the prior in logarithmic scale, so in terms of the relative standard deviation of the original profile variability, with $\sigma_1 = 0.01$, $\sigma_2 = 0.4$, $h_1 = 5$, $h_2 = 27$, $s_1 = 9$, and $s_2 = 6$. The smoothness of the profiles comes from the off-diagonal entries of the prior covariance, and for these a standard Gaussian covariance function is used. The elements of the prior covariance function are thus

$$c_{ij} = \sigma(i)\sigma(j) \exp\left(-\frac{1}{2}\left(\frac{d(i,j)}{\phi}\right)^2\right), \tag{13}$$

where i and j are two altitudes, $d(i,j)$ is the spatial distance between them in kilometers, and the correlation length parameter ϕ controls the smoothness, here we set $\phi = 12$ km. Equation (13) gives rational profiles when we use the positivity condition defined in equation (10). The covariance of equation (13) is shown in Figure 2 (top left). Figure 2 (top right) shows the five first singular values of the decomposed covariance and Figure 2 (bottom left) shows the three largest singular vectors of the original covariance matrix. Any profile suggestion is a linear combination of these three components, weighted by the $\boldsymbol{\alpha}$ parameters. Figure 2 (bottom right) shows 20 random draws using the three largest singular vectors. The random draws visualize typical realizations of equation (10) representing candidate profile shapes for our retrieval attempt.

4.3. Transmittance and Jacobian

In this section we describe more carefully how we model the FTIR spectra. We show the calculation of the Jacobian, which is needed if we use some derivative-based optimization instead of MCMC. Furthermore, the Jacobian is needed later in section 4.4 where we discuss the averaging kernel.

Given the forward model $f(\boldsymbol{x})$, observations $\boldsymbol{y} = [y_1, y_2, \ldots, y_m]^T$, and the state vector $\boldsymbol{x} = [x_1, x_2, \ldots, x_n]^T$, the linearized forward model for some reference state $\boldsymbol{x}_r$ can be written as

$$\boldsymbol{y} - f(\boldsymbol{x}_r) = \frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}}(\boldsymbol{x} - \boldsymbol{x}_r) + \boldsymbol{\varepsilon} = \boldsymbol{K}(\boldsymbol{x} - \boldsymbol{x}_r) + \boldsymbol{\varepsilon}, \tag{14}$$

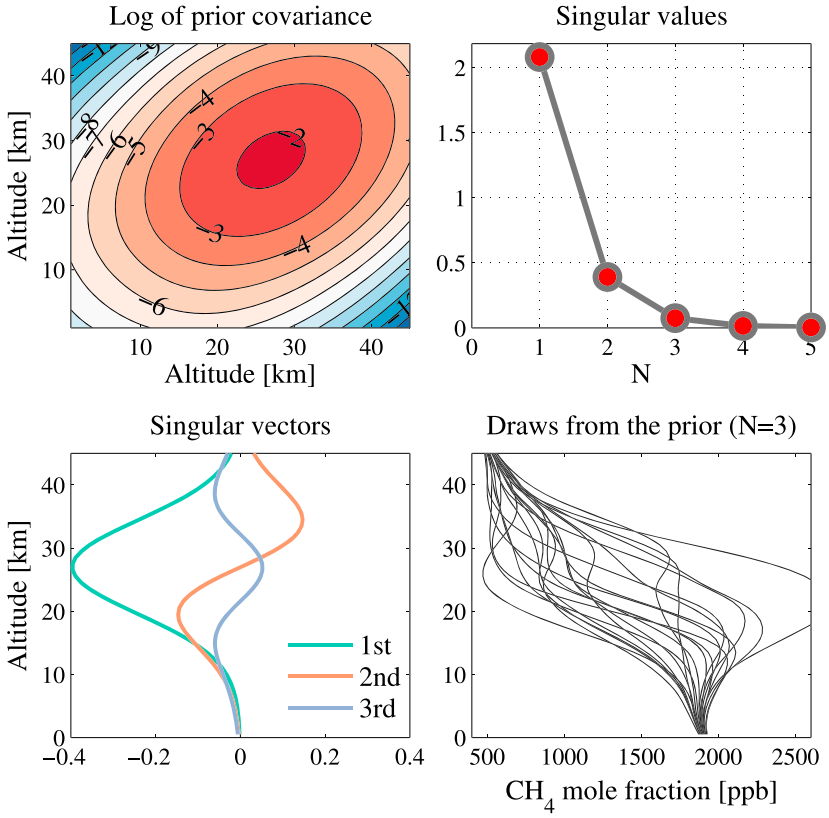

Figure 2. (top left) Logarithm of the CH_4 prior covariance, (top right)five largest singular values, (bottom left) three largest singular vectors, and (bottom right) random draws from the prior using the three singular vectors.

where ε is the vector of measurement errors and $\boldsymbol{K}$ is the $m \times n$ weighting function matrix also called the Jacobian which has the elements $K_{ij} = \partial f_i(\boldsymbol{x})/\partial x_j$. With the direct Sun geometry, the derivatives (for each $\boldsymbol{x}$) can be easily computed analytically. Assuming a discretized atmosphere with n layers, and only one absorbing gas for simplicity, the transmittance in one wavelength is

$$\tau = \exp\left(-\sum_{i=1}^{n} \sigma_i x_i l_i\right), \tag{15}$$

where σ_i is the absorption coefficient, or cross section, of the layer i, and x_i and l_i are the corresponding number density and slant length of the layer. In the matrix notation for the whole spectral window with m points we have the cross-section matrix $\boldsymbol{S} \in \mathbb{R}^{m \times n}$, densities $\boldsymbol{x} \in \mathbb{R}^n$, and slant lengths $\boldsymbol{l} \in \mathbb{R}^n$. Now the slant densities are

$$\bar{\boldsymbol{x}} = \boldsymbol{l} \circ \boldsymbol{x}, \tag{16}$$

where $\circ$ denotes the pointwise product, and the transmittance

$$\boldsymbol{\tau} = \exp(-\boldsymbol{S}\bar{\boldsymbol{x}}), \tag{17}$$

$\boldsymbol{\tau} \in \mathbb{R}^m$, which has the Jacobian

$$\boldsymbol{K} = \frac{\partial \boldsymbol{\tau}}{\partial \bar{\boldsymbol{x}}} = -\mathrm{diag}(\boldsymbol{\tau})\boldsymbol{S}, \tag{18}$$

with $\boldsymbol{K} \in \mathbb{R}^{m \times n}$. However, in the dimension reduction retrieval we estimate $\boldsymbol{\alpha}$ parameters instead of gas densities of the individual layers. Using the logarithmic profile, the projection matrix $\boldsymbol{P}_k$ in equation (10) gives the mapping back to the full space

$$\boldsymbol{x}_k = \exp(\boldsymbol{P}_k \boldsymbol{\alpha}_k) \circ \boldsymbol{x}_0, \tag{19}$$

and the Jacobian becomes

$$\boldsymbol{K}_k = \frac{\partial \boldsymbol{\tau}}{\partial \boldsymbol{\alpha}_k} = \boldsymbol{K}\mathrm{diag}(\boldsymbol{I} \circ \boldsymbol{x}_k)\boldsymbol{P}_k \tag{20}$$

with $\boldsymbol{K}_k \in \mathbb{R}^{m \times k}$. In practice, the situation is slightly more complicated. Instead of measuring transmittance, $\boldsymbol{\tau}$, a ground-based FTS observes solar radiation influenced by the atmosphere and the Fourier transformed spectrum contains a nonphysical baseline (thus, its units are arbitrary). Moreover, an extra additive term called the zero-level offset is usually added to take account the nonlinearity of the detector. Thus, the FTIR spectrum in a narrow wavelength window can be approximated as

$$\boldsymbol{\tau}_{\mathrm{ftir}} = \boldsymbol{\tau} \circ \boldsymbol{s} \circ \boldsymbol{p} + \boldsymbol{\Delta}_0, \tag{21}$$

where $\boldsymbol{s}$ denotes the solar irradiance, $\boldsymbol{p}$ is a polynomial describing the (usually smooth) baseline and $\boldsymbol{\Delta}_0$ is the zero-level offset term. The degree of the polynomial term depends on the width and location of the spectral window; a degree of 1 or 2 is generally feasible. We have used the degree of 2 and parameterized $\boldsymbol{p}$ as a Lagrange polynomial so that the three coefficients are between 0 and 1. The baseline fit should not be constrained; hence, the polynomial coefficients will have flat prior distributions. The offset term is practically always close to zero; a typical value is around 10^{-3}, while $\boldsymbol{\tau}_{\mathrm{ftir}}$ is between 0 and 1. Generally, the offset and the polynomial terms identify much better in wider wavelength windows. The downside of a wide window is that a higher-order polynomial may be required for the baseline.

4.4. Averaging Kernel

It is useful to investigate the resolution and information content of the dimension reduction retrieval. Following *Rodgers* [2000], the averaging kernel is defined as

$$\boldsymbol{A} := \frac{\partial \hat{\boldsymbol{x}}}{\partial \boldsymbol{x}}, \tag{22}$$

which is a measure of the sensitivity of the retrieved state $\hat{\boldsymbol{x}}$ to the change in the underlying truth denoted by $\boldsymbol{x}$. The averaging kernel can be estimated with simulation experiments (when the "truth" is known) or directly from the Jacobian, prior, and related uncertainties. The Jacobian for the vertical profile in full space is

$$\boldsymbol{K}_v = \boldsymbol{K}\mathrm{diag}(\boldsymbol{I}). \tag{23}$$

Now using the formulation by *Rodgers* [2000], and equations (20) and (23), we can write the averaging kernel for the $\boldsymbol{\alpha}_k$-parameters

$$\boldsymbol{A}_\alpha = \frac{\partial \hat{\boldsymbol{\alpha}}_k}{\partial \boldsymbol{x}} = \left(\boldsymbol{K}_k^T \boldsymbol{C}_y^{-1} \boldsymbol{K}_k + \boldsymbol{I}_k\right)^{-1} \left(\boldsymbol{K}_k^T \boldsymbol{C}_y^{-1} \boldsymbol{K}_v\right), \tag{24}$$

where $\boldsymbol{A}_\alpha \in \mathbb{R}^{k \times n}$, $\boldsymbol{C}_y \in \mathbb{R}^{m \times m}$ is the error covariance of the measurement, and $\boldsymbol{I}_k \in \mathbb{R}^{k \times k}$ is the prior covariance of the $\boldsymbol{\alpha}_k$ parameters. We can further write the density-wise averaging kernel matrix

$$\boldsymbol{A} = \mathrm{diag}(\boldsymbol{x}_k)\boldsymbol{P}_k \boldsymbol{A}_\alpha, \tag{25}$$

where $\boldsymbol{A} \in \mathbb{R}^{n \times n}$. Equation (25) is useful when the retrieved FTIR profiles are compared with the reference measurements having much better vertical resolution (such as AirCore). As given in *Rodgers and Connor* [2003], a smoothed version of the reference profile is

$$\boldsymbol{x}_{\mathrm{smooth}} = \boldsymbol{x}_0 + \boldsymbol{A}(\boldsymbol{x}_{\mathrm{high}} - \boldsymbol{x}_0), \tag{26}$$

where $\boldsymbol{x}_0$ is the prior profile and $\boldsymbol{x}_{\mathrm{high}}$ is the high-resolution profile.

The averaging kernel for the total column is

$$\boldsymbol{A}_c = \boldsymbol{I}_v^T \boldsymbol{A} \tag{27}$$

where $\boldsymbol{A}_c \in \mathbb{R}^n$ and $\boldsymbol{I}_v \in \mathbb{R}^n$ contains the lengths of the layers in the vertical direction. Equation (27) describes the sensitivity of the integrated column, it can be used when column values are compared.

Finally, we mention that SWIRLAB also offers possibility to retrieve CH_4 by scaling prior profiles. The averaging kernels for the least squares prior scaling retrieval can be derived in the same way as above and omitted here.

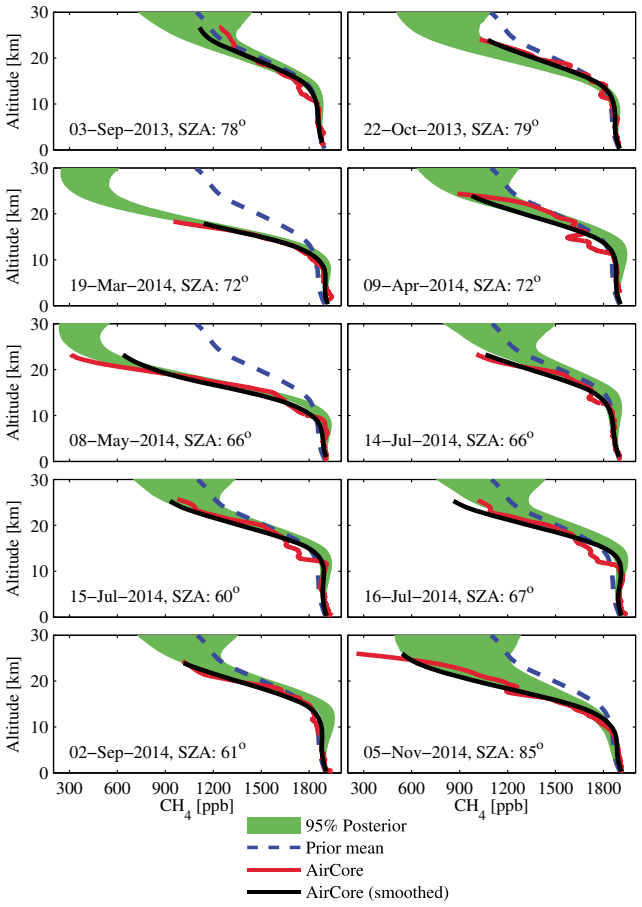

Figure 3. Comparison of the original AirCore in situ profiles (red), smoothed AirCore profiles (black), prior profiles (blue dashed, taken from GGG2014), and the retrieved profiles (green: 95% posterior envelopes of the profiles accepted by the MCMC chain).

4.5. Computation of the Retrieval

Most of the remote sensing retrieval algorithms that apply Bayesian formulation are based on optimization and on the assumption of approximately Gaussian posterior distribution around the maximum a posteriori (MAP) point. Here our aim is to apply a Monte Carlo sampling scheme to calculate the full posterior distribution. In particular, we apply the MCMC methodology, which can be seen as a clever Monte Carlo technique as it avoids the direct computation of the scaling factor needed in the Bayes formula. Note that this factor is not needed when only the MAP point is searched, as, e.g., in the optimal estimation or Levenberg-Marquardt (LM) algorithms.

The MCMC algorithm has previously been applied to atmospheric remote sensing problems by, e.g., *Tamminen and Kyrölä* [2001], *Haario et al.* [2004], and *Laine and Tamminen* [2008]. The advantages of MCMC methodology, in addition to computing the full posterior distribution, include the flexibility to perform the inversion without typical restrictions of Gaussian prior and error distribution. When performing the sampling in high dimensional problems the convergence may be slow and using dimension reduction with the MCMC is a tempting opportunity, which we will demonstrate here. We have used the efficient adaptive MCMC by *Haario et al.* [2001, 2006].

Despite that we have used the adaptive MCMC method and we have just a few parameters to estimate, MCMC method is still substantially slower than derivative-based point estimation methods. The computing time necessary for MCMC depends on the number of samples, number of parameters, used wavelength band, and the computer resources. On a modern desktop PC (Intel Xeon 2.40 GHz), our current CH_4 retrieval takes ~1 min 40 s with 100,000 samples. In comparison, the LM estimation, which we always perform before MCMC to get good initial values, takes just a few seconds. As a Matlab code, SWIRLAB is not optimized for speed, but for user friendliness. It is not intended to be a serious operational processor for large data sets, but rather a tool for research and development. In many cases MCMC and LM give approximately the same results. This is especially true when the error estimate of the FTIR spectrum is correct, which means that the likelihood and the prior are in good balance. If the error estimate is substantially smaller (or larger) than the spectral fit suggests, then the LM results become less reliable. With MCMC, the measurement error needs not to be known beforehand. Our adaptive MCMC algorithm allows estimation of the error during the sampling.

5. Results

In this section we apply the dimension reduction method for the retrieval of CH_4 profile information from FTIR data. We use FTIR measurements made at Sodankylä, Northern Finland, on 10 different days between 3 September 2013 and 5 November 2014. For all of the investigated days, we have a reference in situ measurement made with the AirCore system. The measurements made on 19 March 2014 and 8 May 2014 are

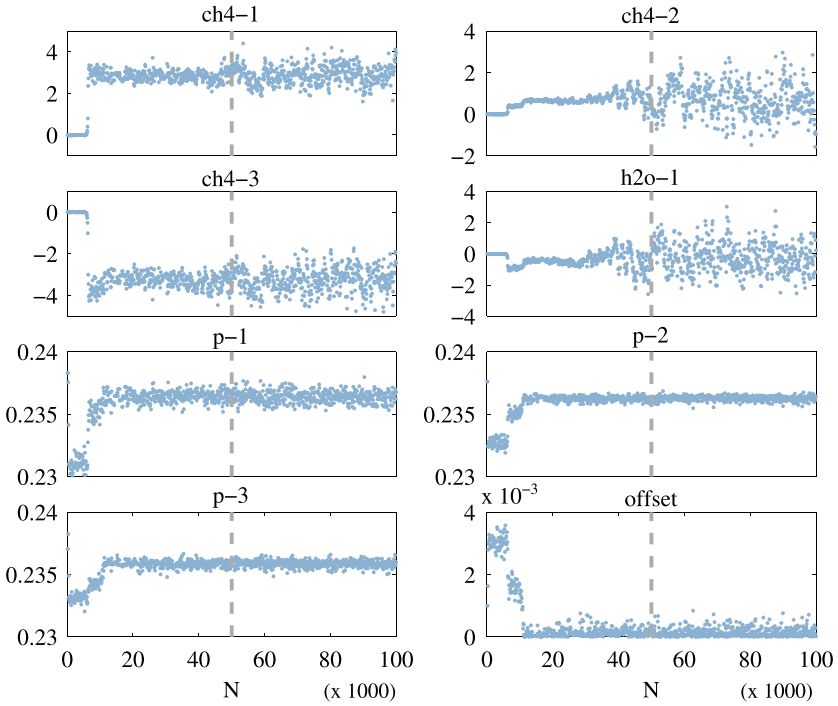

Figure 4. Example of the full MCMC chain of the sampled parameters. Shown are every 100th sample from the chain of 100,000 samples. The dashed grey lines indicate the burn-in periods.

especially good test cases for our retrieval method, because on these two days the measured AirCore CH_4 profile shapes differ substantially from the standard a priori profiles assumed for the stratospheric part of the profiles. This is most probably due to the polar vortex conditions on these days.

Figure 3 shows examples of the retrieved CH_4 posteriors from the 10 investigated days using the covariance of equation (13). The three profile shape parameters were estimated running 100,000 samples of MCMC. The first 50,000 samples were discarded because the MCMC chain takes some time to adapt (this is called the burn-in period). The other retrieved parameters were H_2O prior location (one parameter), zero-level offset (one parameter), and the polynomial baseline (three parameters). The estimated values of these extra parameters are not so relevant for this study and omitted from the discussion. The profile shapes sampled by the MCMC were compared with the original priors (blue) and the AirCore soundings from the same days (red). To smooth the AirCore data using equation (26), the AirCore profiles were extrapolated to the retrieval grid

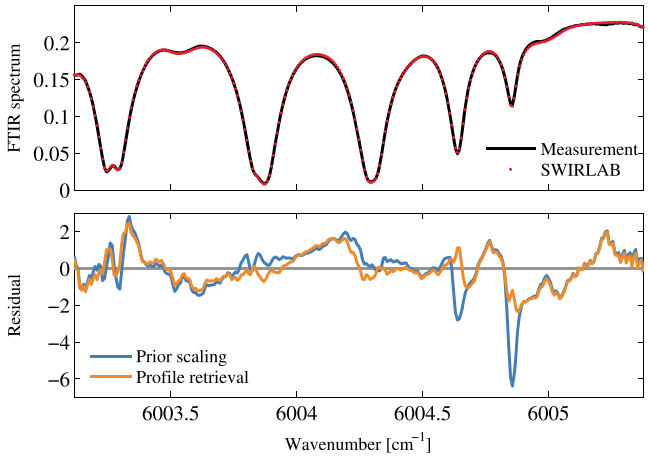

Figure 5. (top) Used wavelength band and an example SWIRLAB fit. (bottom) Comparison of the residuals from the prior profile scaling and dimension reduction methods using data from 19 March 2014.

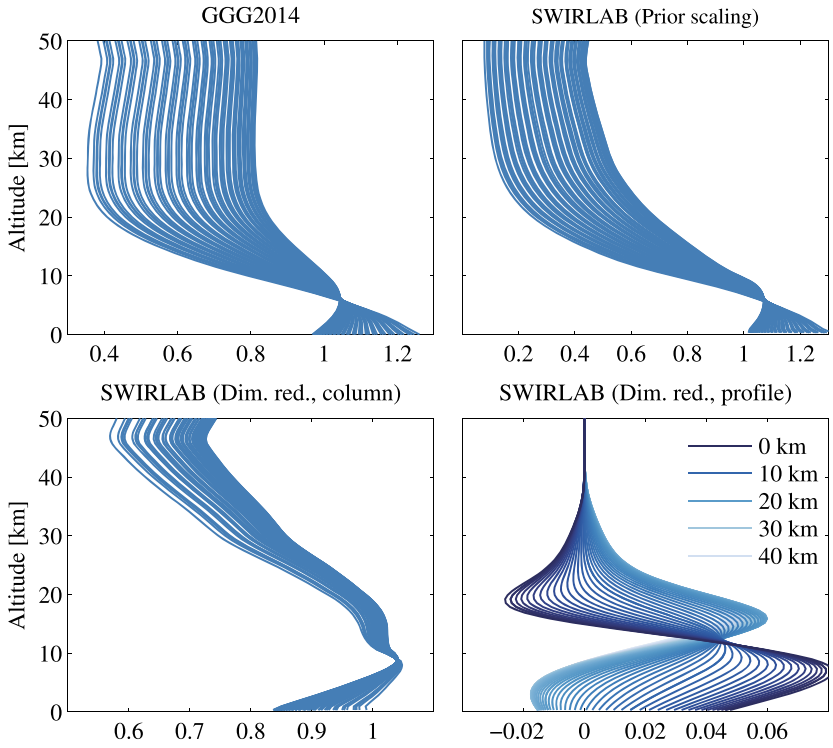

Figure 6. Examples of the CH_4 averaging kernels from 19 March 2014. Shown are averaging kernels (top left) from GGG2014, and (top right) from SWIRLAB using prior profile scaling. Averaging kernels from SWIRLAB using dimension reduction retrieval — (bottom left) for total column and (bottom right) a full averaging kernel matrix.

between 0 and 70 km. Thus, the smoothed AirCore curves (black) are somewhat ambiguous in the upper end of the profiles. We also note that AirCore provides dry-air mole fractions but SWIRLAB uses number density. To convert between units, we used the prior air and H_2O number density profiles. The comparison in Figure 3 is in wet-air mole fraction.

In general, the agreement between the retrieved profiles and the AirCore profiles is very good. In all cases the 95% posterior envelope overlaps with the AirCore profile at almost all altitude levels. With the 19 March 2014 and 8 May 2014 data, our method correctly finds the steep gradient in the CH_4 mole fraction at 10–20 km. In these polar vortex cases, the stratospheric part of the prior profile is clearly incorrect. The MCMC samples that generated the 19 March 2014 profile are shown in Figure 4. The three CH_4 parameters drift relatively far from the initial point, which suggests strong data-based evidence of the profile information. Note that there is a positivity condition in the offset term.

In the spectral fitting, we use a narrow wavelength band between 6003 and 6005.5 cm^{-1} (Figure 5, top). The information about the CH_4 gradient mainly comes from the two medium-strong lines centered at 6004.65 and 6004.86 cm^{-1}. With an incorrect prior profile shape, the prior scaling method produces large residuals in these two lines, a discrepancy that can be improved by fitting three parameters for the profile shape (Figure 5, bottom).

Examples of the averaging kernels from 19 March 2014, defined in equations (25) and (27), are shown in Figure 6. For comparison, we also show the averaging kernels from the closest CH_4 window used by GGG2014 and averaging kernels from the SWIRLAB version that scales prior profiles (GGG-type retrieval). Although not exactly from the same wavelength window, the GGG2014 averaging kernels (Figure 6, top left) and the corresponding SWIRLAB averaging kernels (Figure 6, top right) are visually very similar. The averaging kernels of the dimension reduction retrieval for the total column (Figure 6, bottom left) have a slightly different shape and smaller variation during the day. In Figure 6 (bottom right) is one example of the full averaging kernel matrix.

To illustrate potential errors in the TCCON XCH_4 when the prior profile has incorrect shape, we show a comparison between TCCON, SWIRLAB, and AirCore XCH_4 columns for 19 March 2014 (Figure 7). The Sodankylä

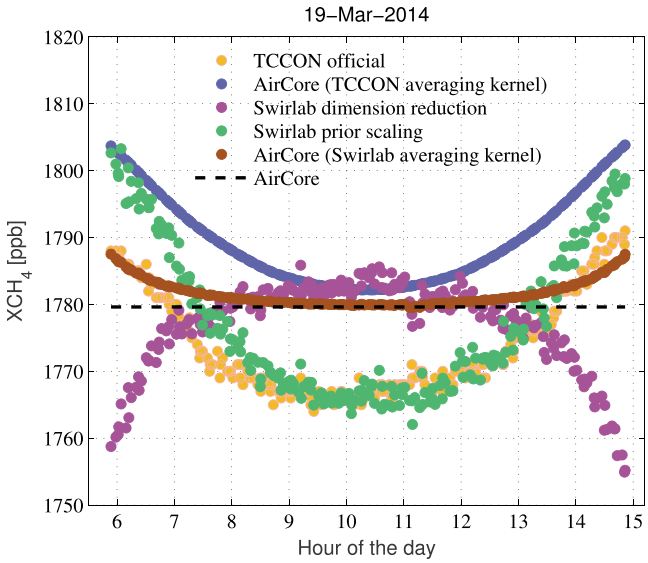

Figure 7. Comparison of the XCH_4 columns derived from FTIR and AirCore. Shown are the official TCCON (orange) and corresponding AirCore (blue), the SWIRLAB dimension reduction (purple), and corresponding AirCore (brown). Also shown are the AirCore column without smoothing (black dashed) and the SWIRLAB prior scaling solution (green).

FTIR data from that day cover solar zenith angles between 67 and 82°. TCCON provides averaging kernels in a 5° solar zenith angle grid; thus, we interpolated them for the solar zenith angles of the measurements. Both AirCore and SWIRLAB XCH_4 were derived according to the equations given in *Wunch et al.* [2011]. It would be better to estimate SWIRLAB XCH_4 using the retrieved O_2, but the current version of SWIRLAB does not yet allow O_2 retrievals. There is a notable "U shape" in the TCCON XCH_4, which is also visible in the smoothed AirCore. There is also a bias of around 16 ppb. Although the TCCON averaging kernels produce similar solar zenith angle dependence to the reference measurements, we suspect that the TCCON XCH_4 values may contain additional errors when the scaled prior has incorrect shape. The solar zenith angle dependence is substantially smaller in the AirCore columns smoothed using the SWIRLAB dimension reduction averaging kernels, as expected (see Figure 6). The XCH_4 values from SWIRLAB (dimension reduction) agree with AirCore within 5 ppb for solar zenith angles less than 75°. There is a notable negative bias up to ~30 ppb with larger solar zenith angles, which is most probably caused by systematic modeling errors remaining in the spectral fit. The slope is even in the different direction than the averaging kernels indicate. Finally, we note that the prior XCH_4 is ~1820 ppb (not shown), quite far from the AirCore value.

6. Discussion

The maximum amount of information can be extracted from the measurements using a full nonlinear retrieval and appropriate a priori information [*Rodgers*, 2000]. Defining an adequate prior is not always an easy task, and is open for criticism. Moreover, special cases like the polar vortex can be difficult to tackle in practice.

When a full nonlinear retrieval is used in optimal estimation, it can lead to problems like instability of the vertical profile or physically unrealistic retrieval [*Senten et al.*, 2012]. Several ideas to circumvent these issues have been developed in the literature; see, e.g., Tikhonov regularization and truncated singular value decomposition discussed in *Rodgers* [2000]. Not surprisingly, the FTIR CH_4 profile retrieval requires substantial regularization. For example, in *Senten et al.* [2012], the authors compare optimal estimation and Tikhonov regularization to IOA, and conclude that all these methods have some problems with the stability (but IOA is less sensitive to the choice of the prior covariance matrix).

The main difference between the dimension reduction approach of this paper and most other approaches is that we solve the problem in low dimensional subspace rather than regularize the profiles in full space. In that sense, the approach used in this paper is closer to methods that use arbitrarily parameterized profile shape [e.g., *Kuai et al.*, 2012; *Yang et al.*, 2010]. The idea of the dimension reduction in atmospheric remote sensing is not new, and, for instance, *Rodgers* [2000] discusses this option also. However, there the idea is to represent the high-resolution profile with *some* linear representation and then to use, e.g., truncated singular value

decomposition to solve the problem. It is left open how the actual projection to the lower dimension is created. In our approach (based on *Marzouk and Najm* [2009] and *Solonen et al.* [2016]), the singular value decomposition is applied to the prior covariance matrix, and the retrieval is constrained to the subspace spanned by these singular vectors.

7. Conclusions

In this paper we used the dimension reduction inversion method for the retrieval of CH_4 profiles. Similar singular value truncation techniques have been used for decades to solve various ill-posed inversion problems. In our method we formulate the prior constraints so that the prior covariance reflects the information content of the observations and it can be expressed with a few significant principal components. Ideally, the number of these components (parameters) matches the number of degrees of freedom in the measurement. This approach is more flexible than the retrieval in a dense vertical grid with strong regularization, which is necessary when the vertical information content is modest in the measured signal but the number of unknowns in the state vector is large. The retrieval by the dimension reduction is not much constrained by the prior profile itself, its covariance just defines an informed subspace where smooth deviations from the prior profile are sought. In addition, by retrieving only a few parameters the algorithms become efficient to compute.

We generated the prior covariance using general assumptions about the variability of atmospheric methane. The tropospheric part of the profile is generally quite well mixed, and stronger gradients are more common in the lower stratosphere. In the future, a large array of AirCore measurements would provide a good database for the construction of a more realistic prior covariance. We used the ACE-FTS satellite data to validate our prior assumption of the natural variability for altitudes between 10 and 30 km and found our constructed covariance reasonable. A more extensive characterization of the prior space would be a useful continuation study.

In this study, we retrieved three parameters to describe the shape of the underlying CH_4 profile. The number of parameters depend on the prior covariance and the complexity of the proposal profiles one wishes to generate. More complex prior covariance requires more parameters to represent it. However, with SWIRLAB we can not currently aim for more complex representation of the profile. Although the Sodankylä FTIR data have excellent spectral resolution and signal to noise ratio, the modeling errors in the spectral fit prevent more information to be retrieved reliably. Missing constituents, spectroscopy errors, and uncertainty in the instrument line shape are common factors that complicate the spectral fit, and typically produce correlated residuals. With a too ambitious retrieval scheme, there is a danger of overfitting. Nevertheless, our analysis shows that there is certainly more than one piece of information about the vertical shape of CH_4 in the FTIR spectra. This information can be extracted even with our simplified forward model, at least when the solar zenith angle is less than 75°. Inversion methods that scale fixed prior profile shape do not fully take advantage of all available information in the spectral data.

We used MCMC for the estimation of the $\boldsymbol{\alpha}_k$ parameters to produce posterior analysis of the CH_4 profile. In contrast to using derivative-based optimization and linearized Gaussian uncertainty assumptions, MCMC samples from the true posterior of the solution. MCMC eventually finds the global optimum regardless of the starting point and returns a set of "acceptable" profiles around the optimum. This leads to more realistic uncertainty estimates for the profile shape and further for the total column. MCMC is usable here because we use only a small wavelength window making the forward model extremely fast to compute. In our setup, it takes a few minutes to estimate one profile with 100,000 samples. The $\boldsymbol{\alpha}_k$ parameters can be estimated (with a few iterations) using the local gradient of the forward model, which could be done if the computation time is a priority.

The AirCore system provides valuable data for validating the retrieved profiles. As shown in Figure 3, the agreement between the retrieved posteriors and the AirCore profiles is generally very good. AirCore profiles can be also used in the validation of XCH_4, although the extrapolation of the AirCore profiles causes uncertainty of around 5 ppb. At least in Sodankylä, the springtime TCCON XCH_4 values often have a substantial dependence on the solar zenith angle — the values are smaller at noon than in the morning and evening. Because the averaging kernels of the prior scaling retrievals are very solar zenith angle dependent, the retrieved columns have dependency when the prior is far from the truth. This is not a major problem in comparisons, if the averaging kernel is used, but it is troublesome in data analyses that directly use the column values. Moreover, an incorrect prior profile shape causes additional uncertainty in the spectral fit and further in the retrieved column.

Figure 7 shows that in this case the actual XCH_4 values may also contain bias, but most probably the discrepancy depends on many factors such as the forward model, used wavelength band, etc. By solving the true profile shape, we are able to reduce the air mass artifact and bias in the integrated total column. Finally, we note that our retrieval method could be easily adapted for satellite measurements also. The significant reduction of the state vector, i.e., the number of estimated parameters, would make it possible to use MCMC in the OCO 2 CO_2 profile retrieval, for example.

Acknowledgments
This study was funded by the Finnish Meteorological Institute and the Academy of Finland (INQUIRE and CARB-ARC projects). The authors would like to thank the ACE team for sharing their data (available at http://www.ace.uwaterloo.ca/public.html). The SWIRLAB code is available at http://github.com/tukiains/swirlab/ and the MCMC toolbox at http://helios.fmi.fi/~lainema/mcmc/.

References

Aires, F., W. B. Rossow, N. A. Scott, and A. Chdin (2002), Remote sensing from the infrared atmospheric sounding interferometer instrument 1. Compression, denoising, and first-guess retrieval algorithms, *J. Geophys. Res. Atmos.*, *107*(D22), 4619, doi:10.1029/2001JD000955.

Aumann, H., et al. (2003), AIRS/AMSU/HSB on the Aqua mission: Design, science objectives, data products, and processing systems, *IEEE Trans. Geosci. Remote Sens.*, *41*(2), 253–264, doi:10.1109/TGRS.2002.808356.

Beer, R. (2006), TES on the Aura mission: Scientific objectives, measurements, and analysis overview, *IEEE Trans. Geosci. Remote Sens.*, *44*(5), 1102–1105, doi:10.1109/TGRS.2005.863716.

Bernath, P. F., et al. (2005), Atmospheric Chemistry Experiment (ACE): Mission overview, *Geophys. Res. Lett.*, *32*(15), L15S01, doi:10.1029/2005GL022386.

Brasseur, G. P., and S. Solomon (2005), *Aeronomy of the Middle Atmosphere*, 3rd revised and enlarged ed., Springer, Dordrecht.

Buchwitz, M., et al. (2006), Atmospheric carbon gases retrieved from SCIAMACHY by WFM-DOAS: Version 0.5 CO and CH_4 and impact of calibration improvements on CO_2 retrieval, *Atmos. Chem. Phys.*, *6*(9), 2727–2751, doi:10.5194/acp-6-2727-2006.

Butz, A., et al. (2011), Toward accurate CO_2 and CH_4 observations from GOSAT, *Geophys. Res. Lett.*, *38*(14), L14812, doi:10.1029/2011GL047888.

Cui, T., J. Martin, Y. M. Marzouk, A. Solonen, and A. Spantini (2014), Likelihood-informed dimension reduction for nonlinear inverse problems, *Inverse Prob.*, *114015*(11), doi:10.1088/0266-5611/30/11/114015.

Devi, V. M., B. D. Chris, L. R. Brown, C. E. Miller, and R. A. Toth (2007), Line mixing and speed dependence in CO_2 at 6227.9 cm^{-1}: Constrained multispectrum analysis of intensities and line shapes in the 30013←00001 band, *J. Mol. Spectrosc.*, *245*(1), 52–80, doi:10.1016/j.jms.2007.05.015.

Dils, B., et al. (2014), The Greenhouse Gas Climate Change Initiative (GHG-CCI): Comparative validation of GHG-CCI SCIAMACHY/ENVISAT and TANSO-FTS/GOSAT CO_2 and CH_4 retrieval algorithm products with measurements from the TCCON, *Atmos. Meas. Tech.*, *7*(6), 1723–1744, doi:10.5194/amt-7-1723-2014.

Doicu, A., S. Hilgers, A. von Bargen, A. Rozanov, K.-U. Eichmann, C. von Savigny, and J. Burrows (2007), Information operator approach and iterative regularization methods for atmospheric remote sensing, *J. Quant. Spectrosc. Radiat. Transfer*, *103*(2), 340–350, doi:10.1016/j.jqsrt.2006.05.002.

Feldman, D. R., W. D. Collins, P. J. Gero, M. S. Torn, E. J. Mlawer, and T. R. Shippert (2015), Observational determination of surface radiative forcing by CO_2 from 2000 to 2010, *Nature*, doi:10.1038/nature14240.

Fischer, H., et al. (2008), MIPAS: An instrument for atmospheric and climate research, *Atmos. Chem. Phys.*, *8*(8), 2151–2188, doi:10.5194/acp-8-2151-2008.

Frankenberg, C., I. ben, P. Bergamaschi, E. J. Dlugokencky, R. van Hees, S. Houweling, P. van der Meer, R. Snel, and P. Tol (2011), Global column-averaged methane mixing ratios from 2003 to 2009 as derived from SCIAMACHY: Trends and variability, *J. Geophys. Res. Atmos.*, *116*, D04302, doi:10.1029/2010JD014849.

Goody, R. M., and Y. L. Yung (1995), *Atmospheric Radiation: Theoretical Basis*, 2nd ed., Oxford Univ. Press, New York.

Gottwald, M., and H. Bovensmann (2011), SCIAMACHY — Exploring the changing Earth's atmosphere.

Haario, H., E. Saksman, and J. Tamminen (2001), An adaptive Metropolis algorithm, *Bernoulli*, *7*(2), 223–242.

Haario, H., M. Laine, M. Lehtinen, E. Saksman, and J. Tamminen (2004), MCMC methods for high dimensional inversion in remote sensing, *J. R. Stat. Soc. Ser. B*, *66*, 591–607.

Haario, H., M. Laine, A. Mira, and E. Saksman (2006), DRAM: Efficient adaptive MCMC, *Stat. Comput.*, *16*(4), 339–354, doi:10.1007/s11222-006-9438-0.

Hase, F., J. Hannigan, M. Coffey, A. Goldman, M. Höpfner, N. Jones, C. Rinsland, and S. Wood (2004), Intercomparison of retrieval codes used for the analysis of high-resolution, ground-based FTIR measurements, *J. Quant. Spectrosc. Radiat. Transfer*, *87*(1), 25–52, doi:10.1016/j.jqsrt.2003.12.008.

Intergovernmental Panel on Climate Change (2013), *Climate Change 2013, The Scientific Basis*, Cambridge Univ. Press, Cambridge, U. K.

Jacquinet-Husson, N. et al. (2011), The 2009 edition of the GEISA spectroscopic database, *J. Quant. Spectrosc. Radiat. Transfer*, *112*(15), 2395–2445, doi:10.1016/j.jqsrt.2011.06.004.

Karion, A., C. Sweeney, P. Tans, and T. Newberger (2010), AirCore: An innovative atmospheric sampling system, *J. Atmos. Oceanic Technol.*, *27*(11), 1839–1853, doi:10.1175/2010JTECHA1448.1.

Kivi, R., and P. Heikkinen (2016), Fourier transform spectrometer measurements of column co_2 at sodankylä, Finland, *Geosci. Instrum. Methods Data Syst.*, *5*(2), 271–279, doi:10.5194/gi-5-271-2016.

Kuai, L., D. Wunch, R.-L. Shia, B. Connor, C. Miller, and Y. Yung (2012), Vertically constrained CO_2 retrievals from TCCON measurements, *J. Quant. Spectrosc. Radiat. Transfer*, *113*(14), 1753–1761, doi:10.1016/j.jqsrt.2012.04.024.

Kuai, L., et al. (2013), Profiling tropospheric CO_2 using Aura TES and TCCON instruments, *Atmos. Meas. Tech.*, *6*(1), 63–79, doi:10.5194/amt-6-63-2013.

Laine, M., and J. Tamminen (2008), Aerosol model selection and uncertainty modelling by adaptive MCMC technique, *Atmos. Chem. Phys.*, *8*(24), 7697–7707, doi:10.5194/acp-8-7697-2008.

Laraia, A. L., R. R. Gamache, J. Lamouroux, I. E. Gordon, and L. S. Rothman (2011), Total internal partition sums to support planetary remote sensing, *Icarus*, *215*(1), 391–400, doi:10.1016/j.icarus.2011.06.004.

Liou, K. N. (2002), *An Introduction to Atmospheric Radiation*, Elsevier, New York.

Marzouk, Y. M., and H. N. Najm (2009), Dimensionality reduction and polynomial chaos acceleration of Bayesian inference in inverse problems, *J. Comput. Phys.*, *228*(6), 1862–1902, doi:10.1016/j.jcp.2008.11.024.

Masiello, G., C. Serio, and P. Antonelli (2012), Inversion for atmospheric thermodynamical parameters of Iasi data in the principal components space, *Q. J. R. Meteorol. Soc.*, *138*(662), 103–117, doi:10.1002/qj.909.

O'Dell, C. W., et al. (2012), The ACOS CO_2 retrieval algorithm—Part 1: Description and validation against synthetic observations, *Atmos. Meas. Tech.*, *5*(1), 99–121, doi:10.5194/amt-5-99-2012.

Predoi-Cross, A., A. R. W. McKellar, D. Chris Benner, V. Malathy Devi, R. R. Gamache, C. E. Miller, R. A. Toth, and L. R. Brown (2009), Temperature dependences for air-broadened Lorentz half-width and pressure shift coefficients in the 30013←00001 and 30012←00001 bands of CO_2 near 1600 nm, *Can. J. Phys.*, *87*(5), 517–535, doi:10.1139/P08-137.

Rodgers, C. D. (2000), *Inverse Methods for Atmospheric Sounding: Theory and Practice*, World Sci., Singapore.

Rodgers, C. D., and B. J. Connor (2003), Intercomparison of remote sounding instruments, *J. Geophys. Res.*, *108*(D3), doi:10.1029/2002JD002299.

Rothman, L., et al. (2013), The HITRAN2012 molecular spectroscopic database, *J. Quant. Spectrosc. Radiat. Transfer*, *130*(0), 4–50, doi:10.1016/j.jqsrt.2013.07.002.

Rothman, L. S., et al. (1998), The HITRAN molecular spectroscopic database and HAWKS (HITRAN atmospheric workstation): 1996 edition, *J. Quant. Spectrosc. Radiat. Transfer*, *60*(5), 665–710.

Schneising, O., M. Buchwitz, M. Reuter, J. Heymann, H. Bovensmann, and J. P. Burrows (2011), Long-term analysis of carbon dioxide and methane column-averaged mole fractions retrieved from SCIAMACHY, *Atmos. Chem. Phys.*, *11*(6), 2863–2880, doi:10.5194/acp-11-2863-2011.

Senten, C., M. De Mazière, G. Vanhaelewyn, and C. Vigouroux (2012), Information operator approach applied to the retrieval of the vertical distribution of atmospheric constituents from ground-based high-resolution FTIR measurements, *Atmos. Meas. Tech.*, *5*(1), 161–180, doi:10.5194/amt-5-161-2012.

Sepúlveda, E., et al. (2014), Tropospheric CH_4 signals as observed by NDACC FTIR at globally distributed sites and comparison to GAW surface in situ measurements, *Atmos. Meas. Tech.*, *7*(7), 2337–2360, doi:10.5194/amt-7-2337-2014.

Solonen, A., T. Cui, J. Hakkarainen, and Y. Marzouk (2016), On dimension reduction in Gaussian filters, *Inverse Prob.*, *32*(4), 045003.

Sung, K., L. R. Brown, R. A. Toth, and T. J. Crawford (2009), Fourier transform infrared spectroscopy measurements of H_2O-broadened half-widths of CO_2 at 4.3 μm, *Can. J. Phys.*, *87*(5), 469–484, doi:10.1139/P08-130.

Tamminen, J., and E. Kyrölä (2001), Bayesian solution for nonlinear and non-Gaussian inverse problem by Markov chain Monte Carlo method, *J. Geophys. Res.*, *106*(D13), 14,377–14,390.

Wunch, D., G. C. Toon, J.-F. L. Blavier, R. A. Washenfelder, J. Notholt, B. J. Connor, D. W. T. Griffith, V. Sherlock, and P. O. Wennberg (2011), The total carbon column observing network, *Philos. Trans. R. Soc. London, Ser. A*, *369*(1943), 2087–2112, doi:10.1098/rsta.2010.0240.

Wunch, D., G. C. Toon, V. Sherlock, N. M. Deutscher, X. Liu, D. G. Feist, and P. O. Wennberg (2015), *The Total Carbon Column Observing Network's GGG2014 Data Version*, Carbon Dioxide Information Analysis Center, Oak Ridge Natl. Lab., Oak Ridge, Tenn.

Yang, K., X. Liu, P. K. Bhartia, N. A. Krotkov, S. A. Carn, E. J. Hughes, A. J. Krueger, R. J. D. Spurr, and S. G. Trahan (2010), Direct retrieval of sulfur dioxide amount and altitude from spaceborne hyperspectral UV measurements: Theory and application, *J. Geophys. Res. Atmos.*, *115*, D00L09, doi:10.1029/2010JD013982.

Yoshida, Y., Y. Ota, N. Eguchi, N. Kikuchi, K. Nobuta, H. Tran, I. Morino, and T. Yokota (2011), Retrieval algorithm for CO_2 and CH_4 column abundances from short-wavelength infrared spectral observations by the greenhouse gases observing satellite, *Atmos. Meas. Tech.*, *4*(4), 717–734, doi:10.5194/amt-4-717-2011.